Protein Management

- Chaperone-assisted folding.
- Protein disposal.
  - Ubiquitinylation pathway.
  - Proteasomes / Exosomes.
- Post-translational modification – phosphorylation.

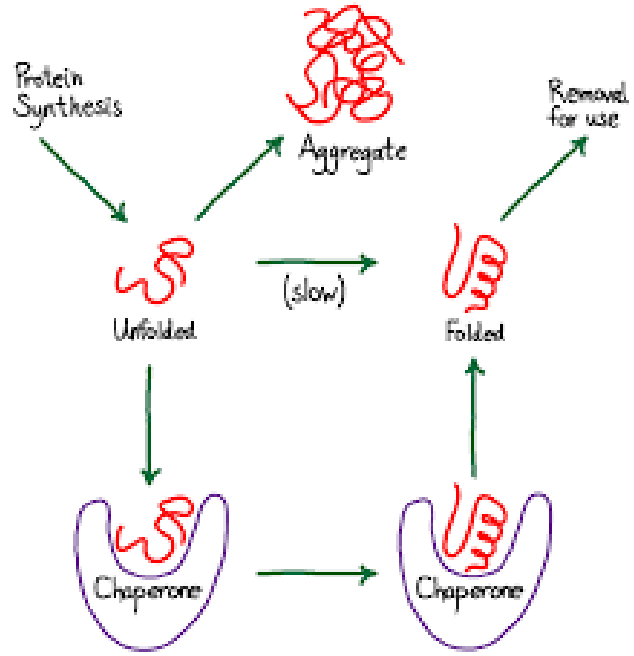
Chaperones: Evolutionary Questions, Mostly Unanswered

- What are the mechanisms by which chaperones recognize their appropriate client proteins?
- Do certain classes of chaperones coevolve with individual client proteins in ways that make them less effective with others?
- Chaperone addiction: once a protein becomes reliant on chaperone assistance for proper folding, does this act as a evolutionary trap by further relaxing selection on features essential to self-folding?
- Given that chaperones consume ATP in the folding process, what is the energetic cost to the cell to producing and relying upon chaperones?
- Does chaperone dependence facilitate the evolution of adaptations that would otherwise not be possible because of their negative effects on self-folding?

Defense Against Protein Misfolding: an ancient intracellular surveillance mechanism.

- Chaperones interact with their client proteins noncovalently, preventing inappropriate aggregation by isolating them.
  - Typically recognize hydrophobic side chains of proteins exposed in their non-native states, then reconfigure them to the native state.

- The investment in protein surveillance is large:
  - On the order of 1% of the total protein in cells can be devoted to chaperones.
  - In prokaryotes, ~25% of proteins are serviced by chaperones; ~10% absolutely depend on them.
  - The numbers are even higher in eukaryotes.
  - Chaperone activity consumes ATP and time.
    - A single engagement typically requires ~5 ATPs, and multiple rounds of engagement are often required to complete the job, which can take on the order of 1 minute per client protein.



- Many viruses rely on host chaperones to assemble capsid proteins to complete their life cycles.

Diverse Sources of Origin of Multiple Molecular Chaperones in Eukaryotes

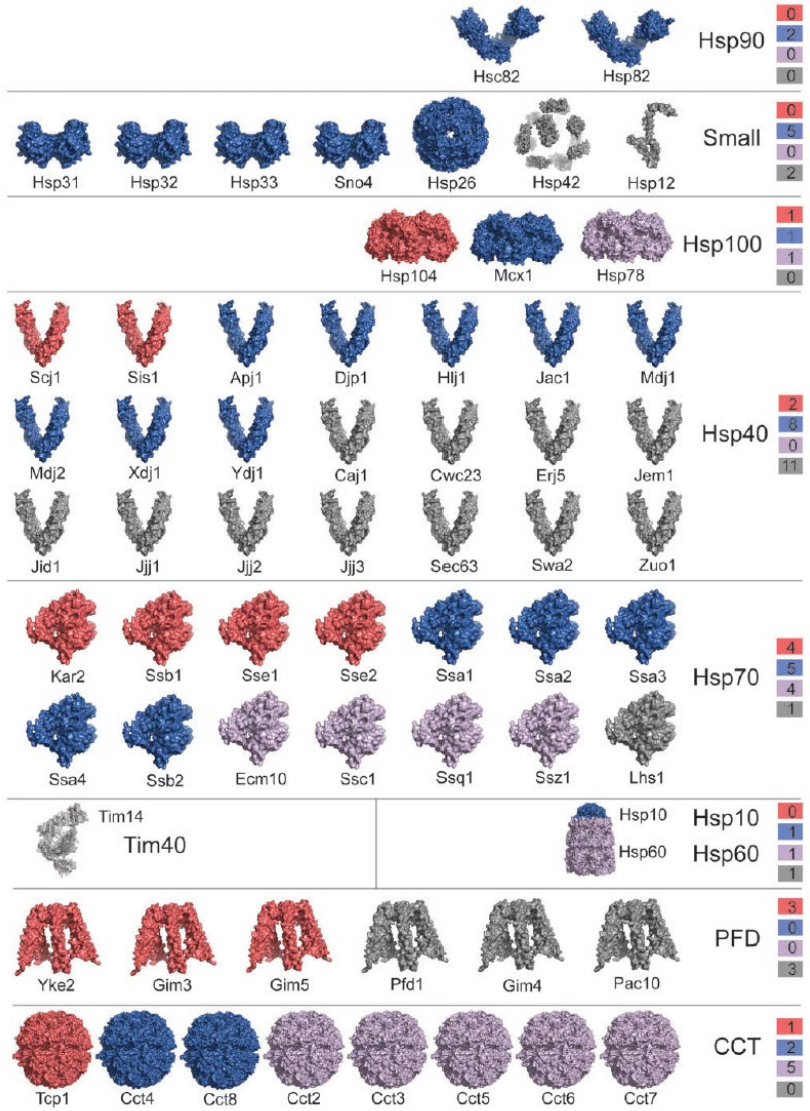


Fig. 1. Yeast chaperones and their reconstructed ancestries. Archaebacterial ancestry is shown in red and eubacterial ancestry in blue. Chaperones with ambiguous ancestry or no homology to prokaryotic proteins are colored in purple and gray, respectively. Here we use the same structural model for all members of the same family. Note that paralogs may deviate in their protein structures. Molecule plots were generated using the PyMOL Molecular Graphics System, version 1.5.0.4 (Schrödinger, LLC).

- Eukaryotic chaperones are derived from archaeal and bacterial sources.

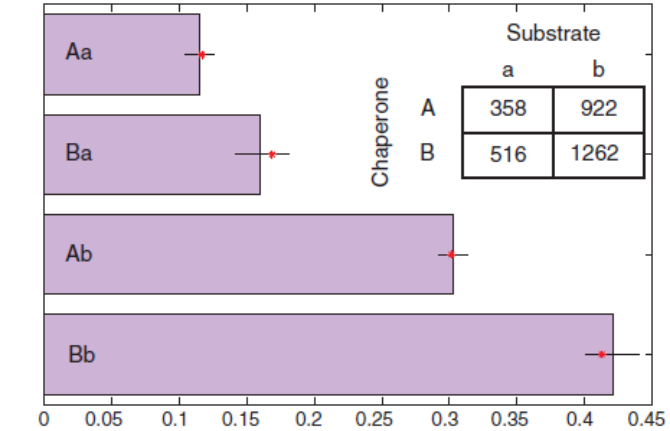


Fig. 2. Prokaryotic origin and connectivity distribution. Asterisks indicate the observed percentage of edges in the network, and bars show the mean expected frequency from randomization simulations. Lines indicate the 1–99 percentile range. Abbreviations: A, archaebacterial, B, eubacterial; uppercase indicates chaperones and lowercase indicates substrates.

- Client genes of the two types are largely unassociated with ancestry.

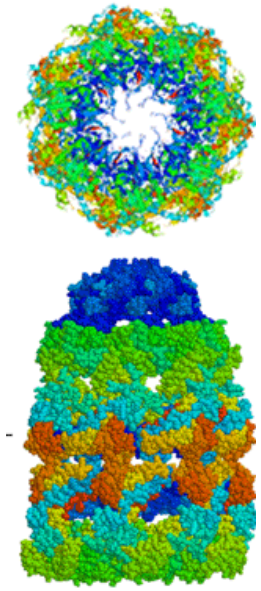
Bogumil et al. (2014, Mol. Biol. Evol.)

The Bacterial Chaperonin Complex

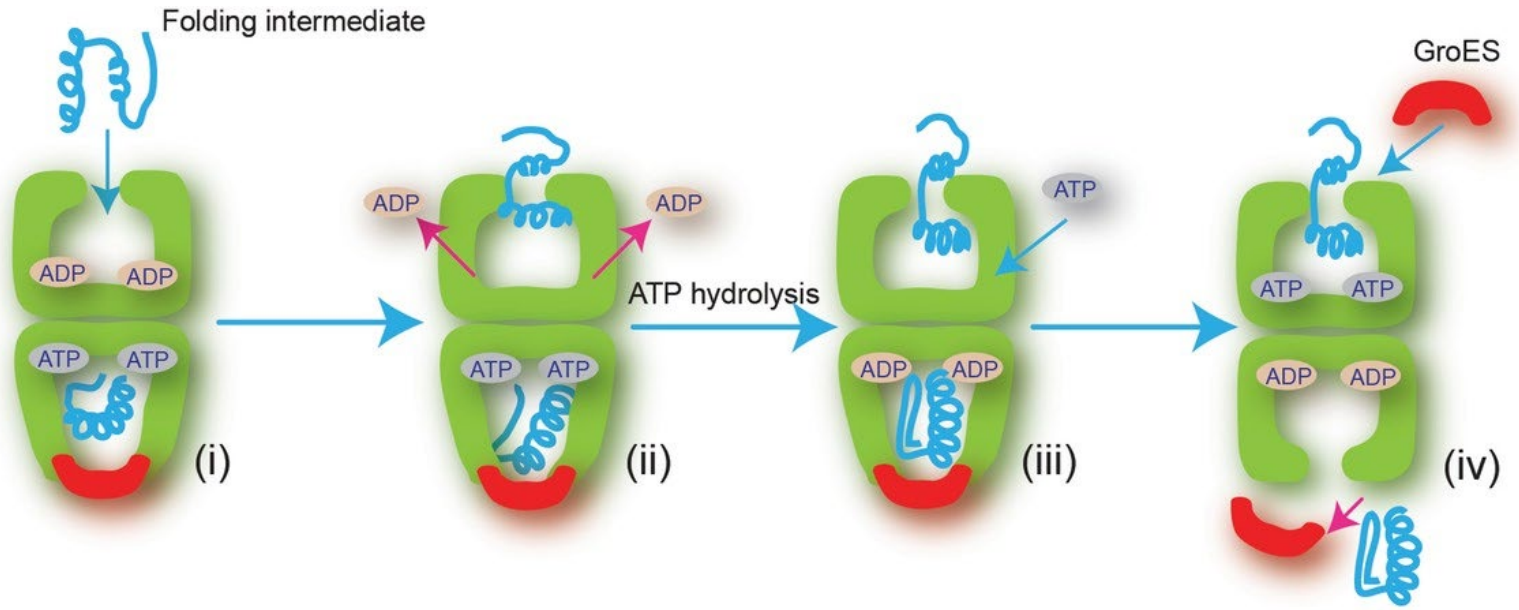
- Most bacteria have a **GroEL chaperonin** system – a homo-tetradecamer (14 subunits, all the same), oligomerized into cages. With the help of GroES, unfolded proteins are moved into the interior for proper folding.

- Eukaryotic descendants: mitochondria and chloroplasts also have GroEL (nuclear encoded).

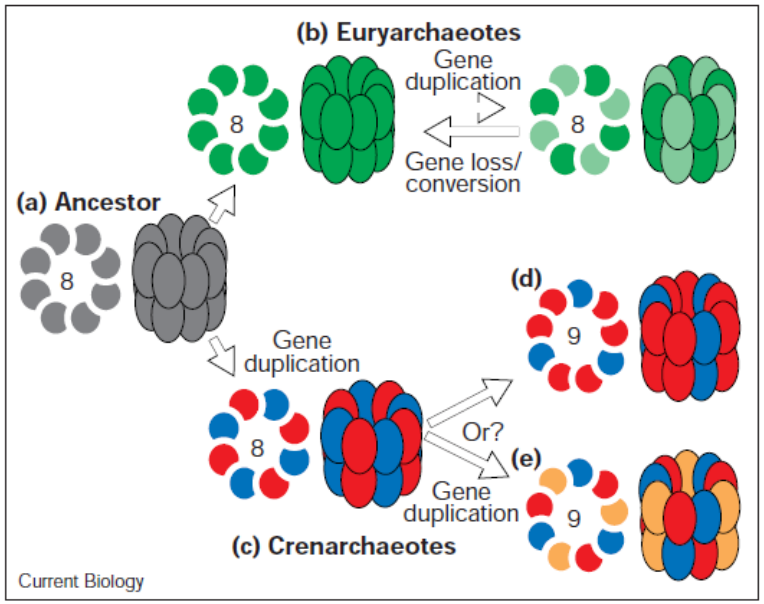
GroES



GroEL



Potential Evolutionary Routes to Alternative Chaperonin Structure



Archaeal chaperonin evolution by recurrent paralogy. Schematic representation of chaperonin structures: multimeric chaperonin rings are composed of individual subunits that interact asymmetrically (side-to-side and top-to-bottom). Subunit colors are the same as in Figure 1; hypothesized interactions between rings are based on *T. acidophilum* [16]. **(a)** Hypothetical ancestral state of the chaperonin complex common to euryarchaeotes, crenarchaeotes and, probably, eukaryotes: eight-membered homo-oligomeric rings (see text). **(b)** Chaperonin subunit gene duplications have occurred independently in at least five euryarchaeal lineages (different subunits are indicated by light and dark green). At least one gene loss has also occurred. **(c)** A gene duplication took place early in crenarchaeal evolution. A more recent gene duplication took place in a *Sulfolobus* ancestor; a change from eight- to nine-membered chaperonin rings also occurred. **(d,e)** Two possible nine-membered structures. (d) The $(\alpha_2\beta)_3$ arrangement of Ellis *et al.* [17] inferred from the two-dimensional crystallization of *Sulfolobus* chaperonins. (e) Our prediction of alternating α, β and γ subunits in each *Sulfolobus* chaperonin ring.

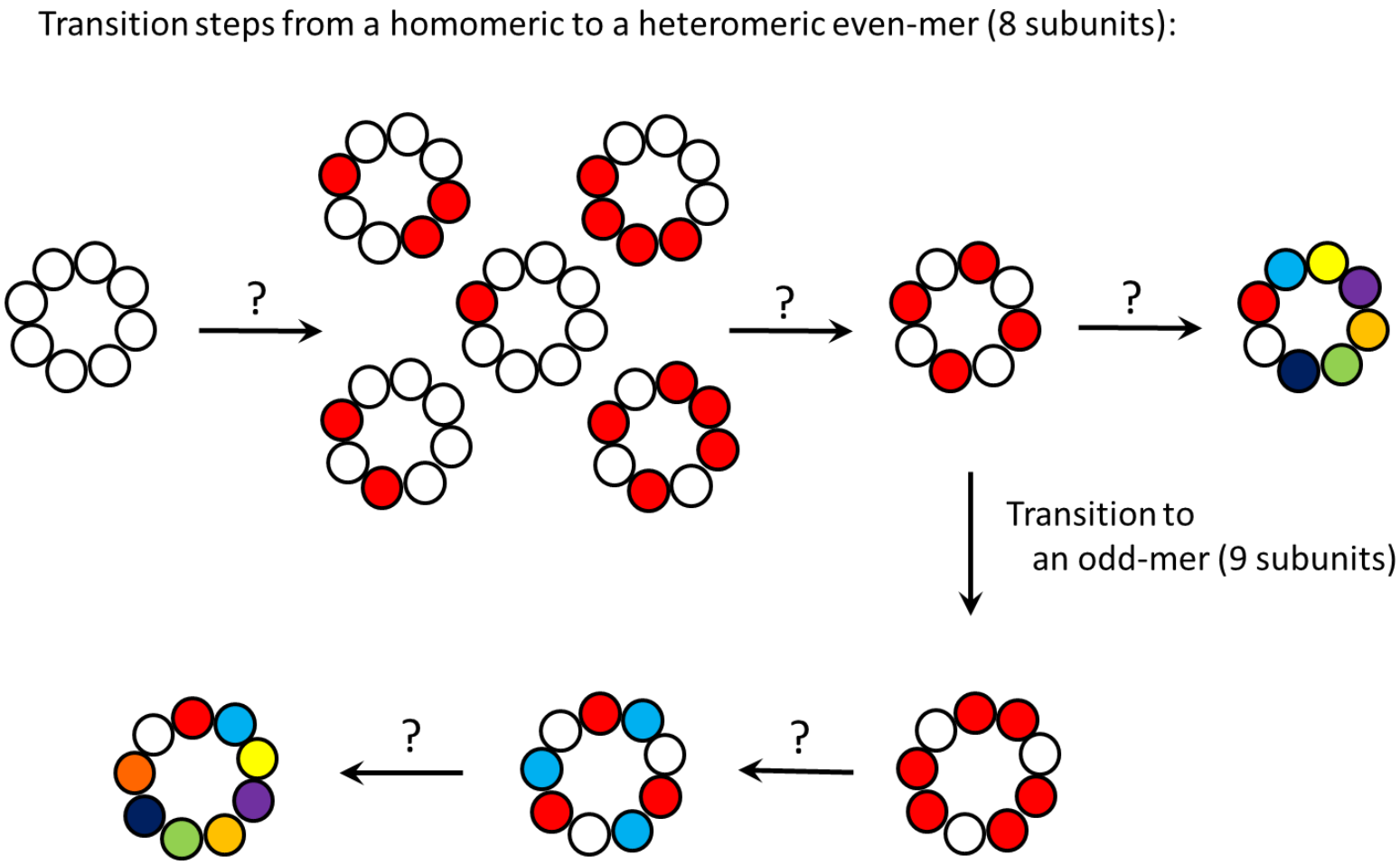
CCT Complex in Archaea and Eukaryotes

- A plausible “semi-neutral” path to heteromer evolution?
- Which came first – subunit-specific roles, or an ordered arrangement of subunits?

Archibald et al. (1999, Curr. Biol.)

How does a heteromeric ring evolve a specific order of subunits?

(rings are also found in the nuclear pore, the proteasome, and some DNA-binding proteins)



- At each “step” in the evolutionary process, at least two mutations would seem to be required.

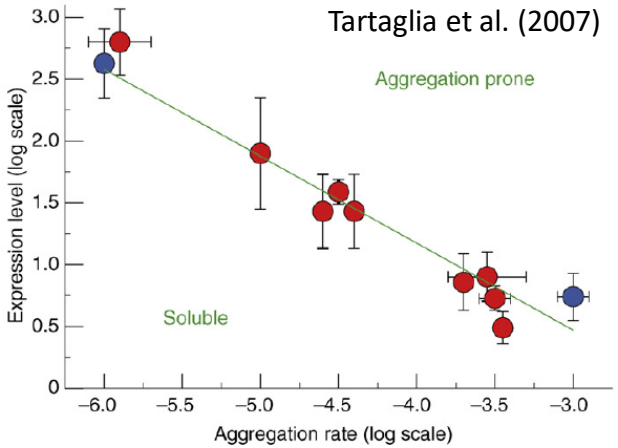
Coevolution of Chaperonins and Their Client Proteins

- Longer proteins are more chaperone dependent.

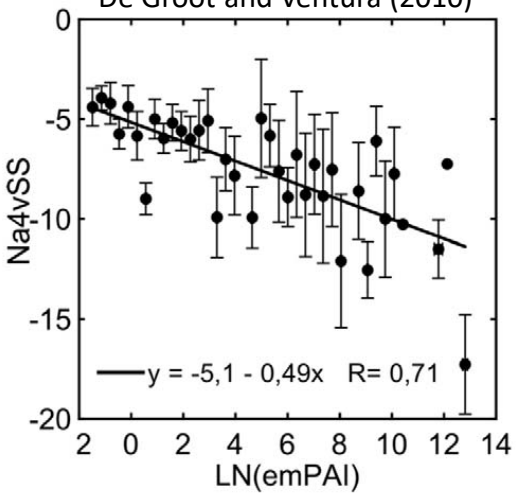
- Inverse relationship between protein expression level and aggregation propensity.

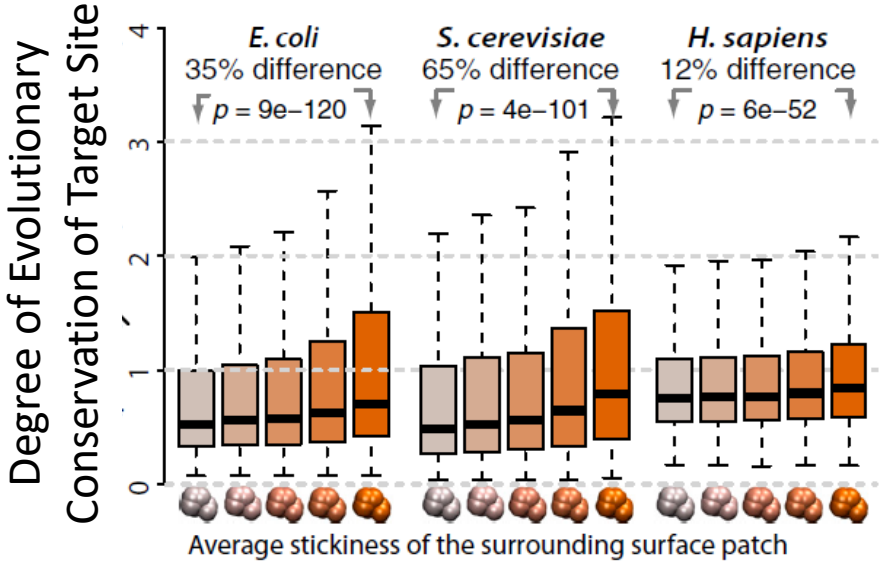
- Amino acids in sticky environments tend to be more evolutionarily conserved, but less so in vertebrates than in microbes.

Tartaglia et al. (2007)



De Groot and Ventura (2010)





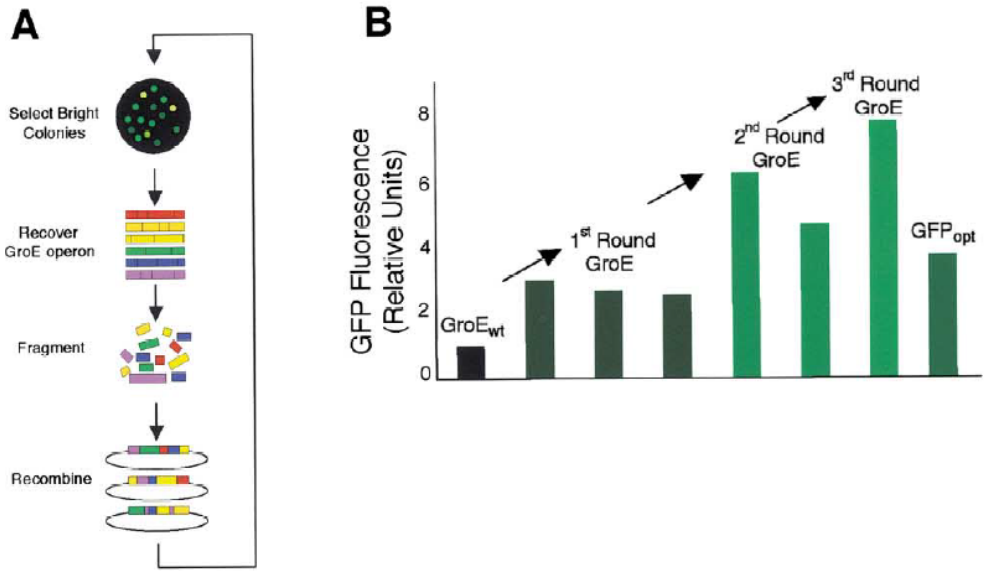
Levy et al. (2012, PNAS)

Evidence for Coevolution of Chamber and Cargo Molecular Motifs

- GroEL can bind and release actin and tubulin (eukaryotic-specific proteins) in an ATP-dependent manner, but does not fold them.
- CCT chaperonin does not even bind some GroEL substrates.

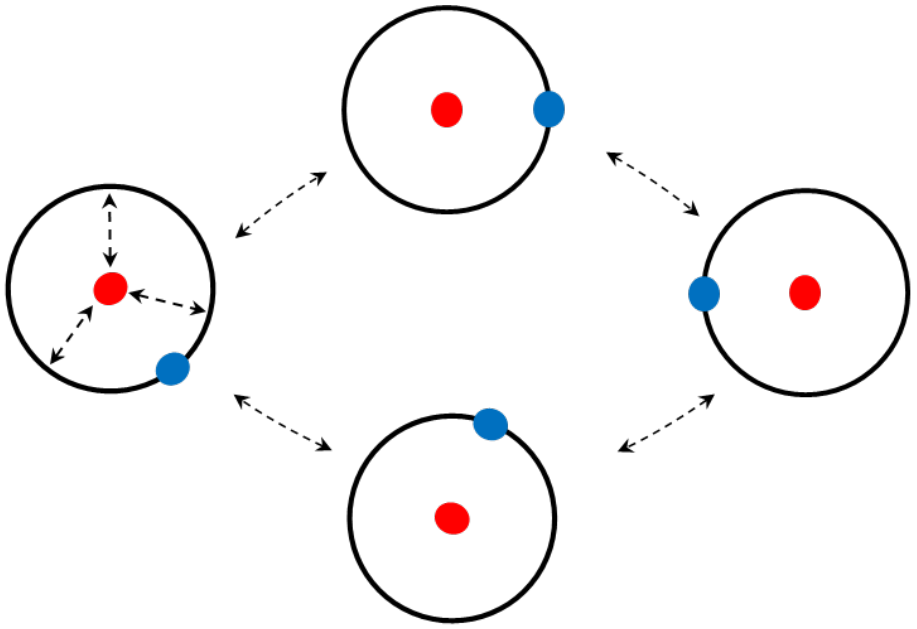
Experimental evolution of GroEL to bind a novel GFP substrate: *in vivo* selection, mutagenic PCR, and *in vitro* DNA shuffling:

- Improved ability to fold GFP came at the expense of the ability to bind natural substrates.



(Wang et al., Cell, 2002)

Coevolutionary Constraints on Client-Protein Recognition: specialists vs. generalists.



- Single client protein: recognition motifs can mutually drift through sequence space, conditional on joint compatibility.

- In principle, chaperone specialization might be achievable by gene duplication, divergence, and expression time/location specialization.

- Multiple client proteins: chaperone motif is expected to become conserved, as improved affinity for one client leads to reduced affinity for others.

Do Proteins Become Evolutionarily Addicted to Chaperones?

- Does chaperone dependence relax selection on client proteins, allowing deleterious misfolding mutations to accumulate?

    - Proteins that are clients of chaperones evolve more rapidly than those that are not (Williams and Fares 2010).

    - In *E. coli*, genes whose protein products are clients of the molecular chaperone GroEL harbor significantly lower frequencies of optimal codons (and hence are expected to experience higher rates of translational error) than do sporadic clients (Warnecke and Hurst 2010).

Deleterious-Mutation Accumulation and Compensation by Elevated Chaperone Levels

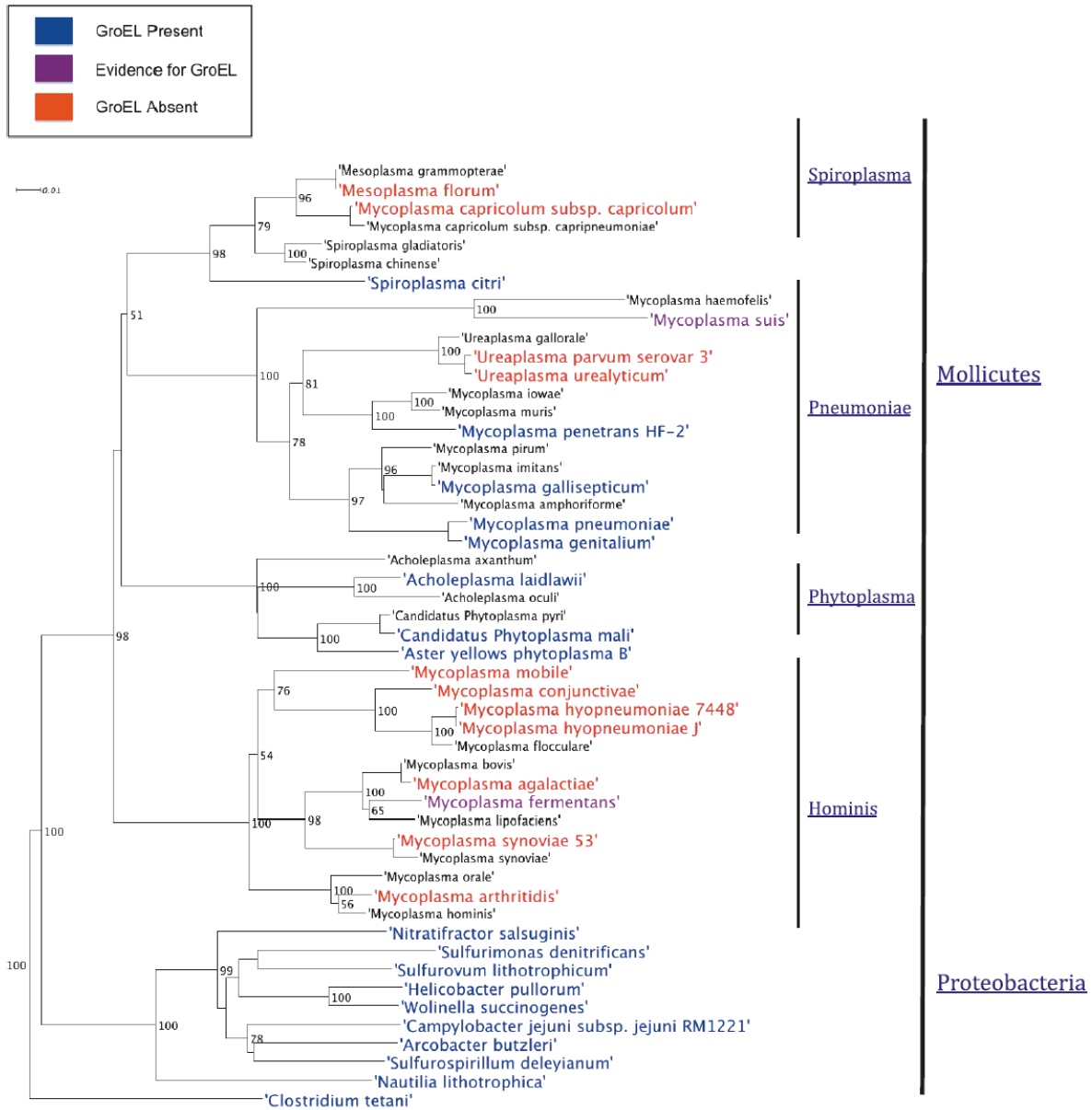
- Experiments with *E. coli* and *S. typhimurium* show that deleterious-mutation accumulation leads to a situation in which survival is enhanced by the overexpression of GroEL or DnaK (Fares et al. 2002; Maisnier-Patin et al. 2005; Van Dyk et al. 1989). This suggests that elevated mutation loads tend to select for genotypes with higher expression of chaperones.

- GroEL is the most highly expressed gene in the aphid endosymbiotic bacterium *Buchnera*, constituting about 10% of total protein (Baumann et al. 1996), and it is also up-regulated in other bacterial endosymbionts (Aksoy 1995; Sato and Ishikawa 1997; Charles et al. 1997; Haines et al. 2002).

    - Moran's (1996) hypothesis: up-regulation of GroEL in endosymbionts is an evolutionary response to accumulated protein-folding problems that arise in species experiencing elevated levels of random genetic drift.

- Several *Mycoplasma* and *Ureaplasma* species (intracellular pathogens) have lost chaperonins entirely.

Fig. 1. 16S rRNA tree from RDP (Cole et al. 2009), created using Weighbour weighted neighbour-joining tree. Bootstrap values greater than 50% are shown. The presence of GroEL is indicated by shading (see inset); black text indicates genomes that are not fully sequenced or lack experimental evidence for the presence of GroEL. Subclades of the *Mycoplasma* are defined (Hominis, Pneumoniae, Spiroplasma, and Phytoplasma).



Clark and Tillier, 2010, Biochem. Cell Biol.

Evolutionary Capacitance: chaperones encourage the maintenance of a reservoir of cryptic adaptive variation?

- Does chaperone dependence facilitate the evolution of adaptations that are otherwise not be possible because of their negative effects on folding?

- "When heat-shock protein buffering is compromised, for example by temperature, cryptic variants are expressed and selection can lead to the continued expression of these traits, even when chaperone function is restored.

  This provides a plausible mechanism for promoting evolutionary change in otherwise entrenched developmental processes."

  (Rutherford and Lindquist 1998).

Figure 1 Developmental abnormalities associated with Hsp90 deficits. See Table 1 for coding of traits. Deformities appearing in *Hsp83* mutant stocks: **a**, *13F3/TM6B*, deformed fore-leg (code L1) and transformed 2nd leg (L2) with an ectopic sex-comb (arrow); **b**, *P582/TM6B*, deformed eye (E1) with an extra antennae (arrow); **c**, *e1D/TM6B*, smooth eyes (E3) with black facets (E5); **d**, *P582/TM6B*, eye margin transformed into scutellum (E2). Abnormal F_1 hybrids produced from crosses between between *Hsp83* mutant stocks and marked laboratory strains: **e**, *e6DX582*^{e1}, left eye has black facets (E5); **f**, *e6AXdpp*^{4}, disorganized abdominal tergites (A1); **g**, *e1DXTM3, ftz-lacz*, small wings (W1); **h**, *e3AXln(2RH)PL,w*^{m}, extraneous tissue growing out of tracheal pit (A2, arrow), **i**, *19F2XCdc37*^{e4D}, eyes absent (E6); **j**, *13F3XCdc37*^{e1E}, wing margin material growing into wing; **m**, *19F2X582*^{e1}, deformed eye. Heteroallelic *Hsp83* combination *e1D/9J1*: **k**, severely deformed legs (L1), **l**, severe black-facet phenotype (E5). Abnormal F_1 hybrids produced with wild-type laboratory stocks and *Hsp83* mutants: **n**, *e1D* or *9J1XIR-6*, thickened wing veins (W3); **o**, *P582XSamarkind*, transformed wing (W5) and extra scutellar bristle (B2, arrow). Abnormalities in wild-type lines raised on geldanamycin: **p**, *IND-6*, notched wings (W2); **q**, *Ore-R*, deformed eye (E1).

A Few Problems

- Because chaperones service hundreds of client proteins, for adaptive capacitance to work, the exposure of any single transiently beneficial variant must outweigh the consequences of a vast array of other exposed deleterious variants.

- The suppression of chaperone activity can lead to the release of mobile-element activity (Specchia et al. 2010) and/or elevated rates of production of aneuploid progeny (Chen et al. 2012), imposing additional negative consequences.

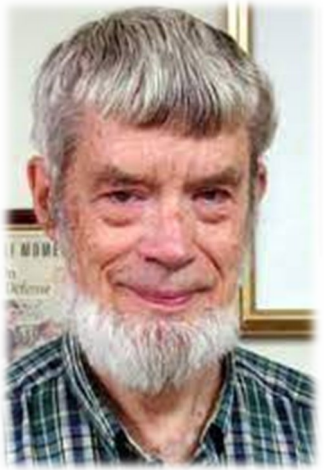
- The argument assumes that chaperones normally buffer the effects of new mutations.

  - In yeast, the phenotypic effects of *de novo* mutations are actually magnified on average (Geiler-Samerotte et al. 2016), implying that natural selection promotes alleles whose effects are less responsive to chaperone limitations, not the other way around.

- For the entire scenario to work, chaperone stress must last long enough to keep the extreme phenotype exposed to selection to enable mutation to produce a mechanism for constitutive expression, but short enough to avoid population extinction.

- The Group-Selection Problem:

  - The proposed narrative relies on the idea that chaperone limitations are promoted on the basis of their long-term benefit to the species, ignoring the fact that individual-level selection operates in the present and has no capacity to see into the future.



George C. Williams
Adaptation and Natural Selection
(1966)

The Proteasome and the Ubiquitin Pathway

- A four-layered barrel in eukaryotes, archaea, and actinobacteria, with each layer being a heptameric ring.
- In archaea, the components are encoded by two genes, alpha forming the catalytically inactive outer layers and beta forming the catalytic core.
- In bacteria, the proteasome is comprised of two layers of hexameric rings, all encoded by a single gene.

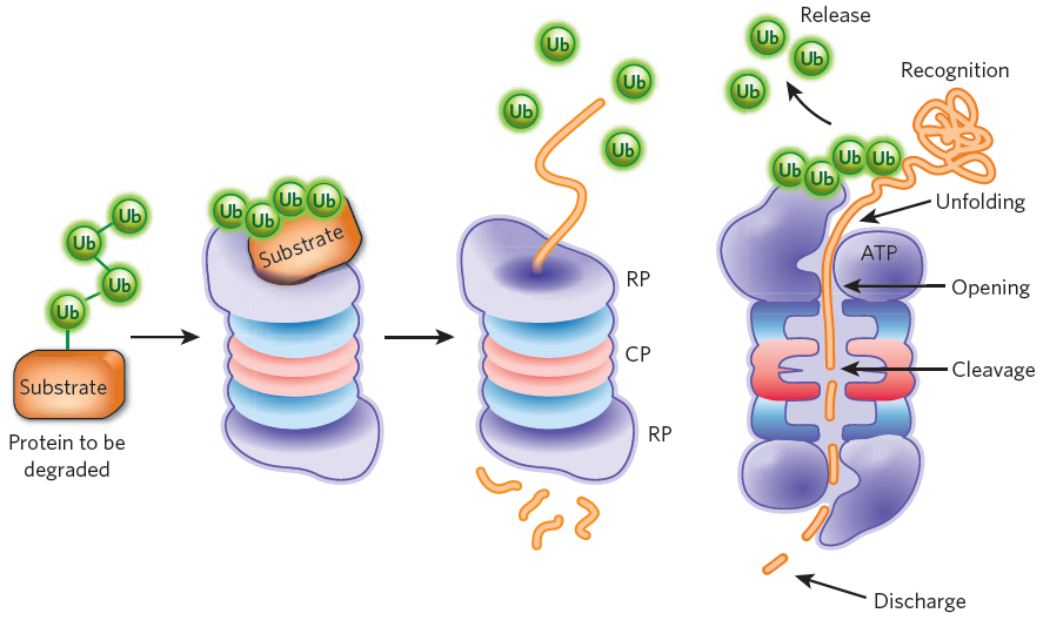
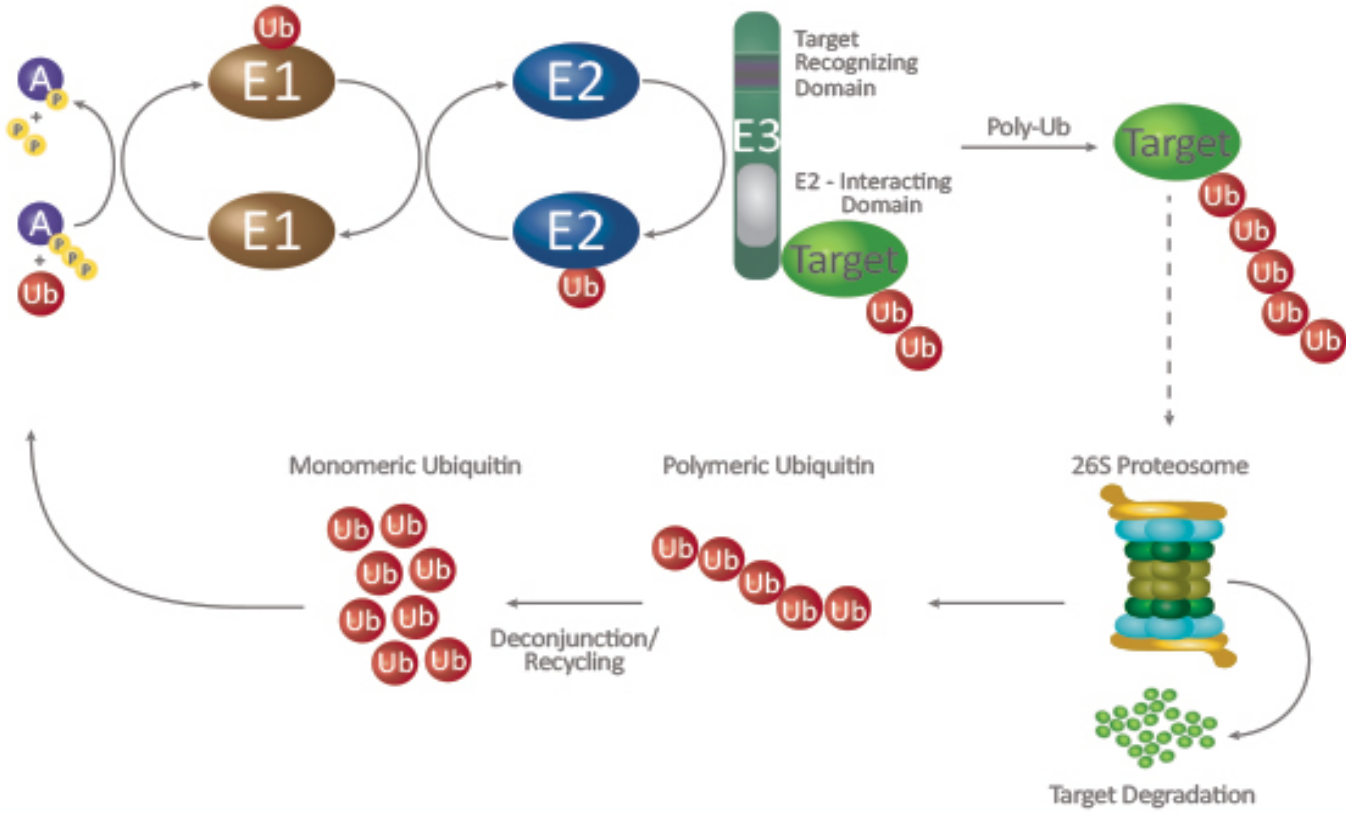


Figure 2 | Polyubiquitin-tagged proteins are often targeted for proteasome-mediated degradation. The ubiquitin-proteasome pathway is responsible for the degradation of hundreds, and probably thousands, of proteins. Many of these substrates are regulatory proteins, such as transcription factors or cell-cycle regulators; others are misfolded or otherwise aberrant proteins that must be eliminated to prevent their aggregation or toxicity. A polyubiquitin-modified protein is the form most commonly targeted to the proteasome. Ubiquitin receptors either in the proteasome regulatory particle (RP, purple) of the 26S proteasome or adaptor proteins that associate reversibly with both polyubiquitylated proteins and specific proteasomal subunits (not shown) allow binding of the proteolytic substrate to the proteasome. As shown in the cut-away on the right, ATPases within the RP unfold the substrate and translocate it into the 20S proteasome core particle (CP, blue and red rings), which houses the proteolytic sites in an interior chamber. The substrate is cleaved to small peptides. Ubiquitin itself is normally recycled by DUBs that bind to or are intrinsic to the RP.

The Ubiquitinylation Pathway



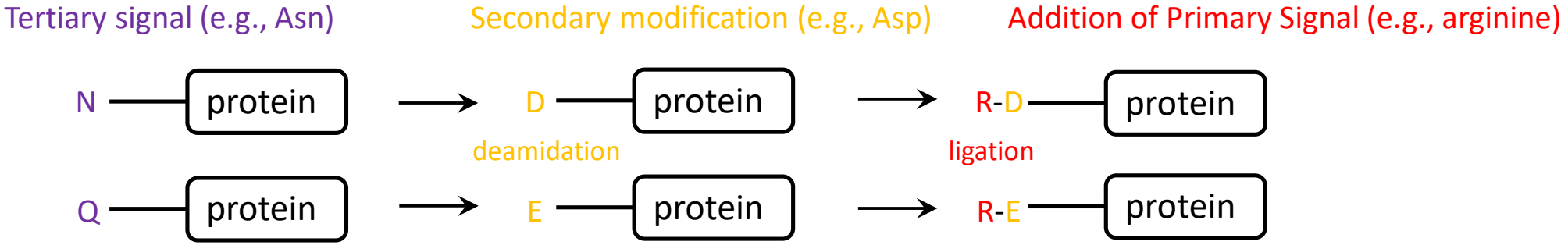
- Through a series of three steps, ubiquitin molecules are addressed to specific target amino acids, and growing chains designate the molecule for destruction.
- Ubiquitins are removed and recycled upon entry to the proteasome.

Some Aspects of the N-end Rule Language

Primary destabilizing and stabilizing residues:

- Defined by the set of destabilizing N-terminal residues in a species.
- Generally has a hierarchical structure.
- Exact mechanisms vary among taxa.

Modification and addition of destabilizing residue:

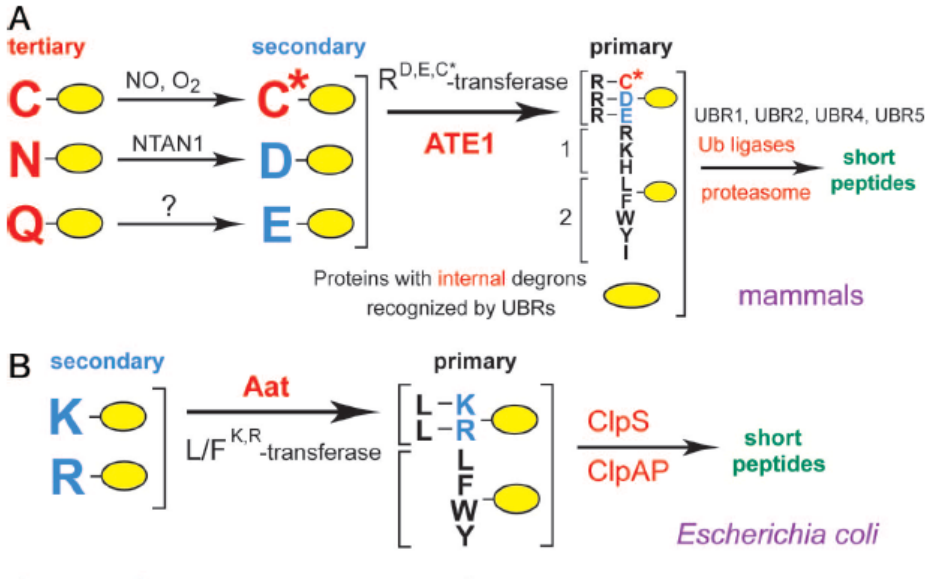


Bacteria Have a Different Set of Rules and Participants

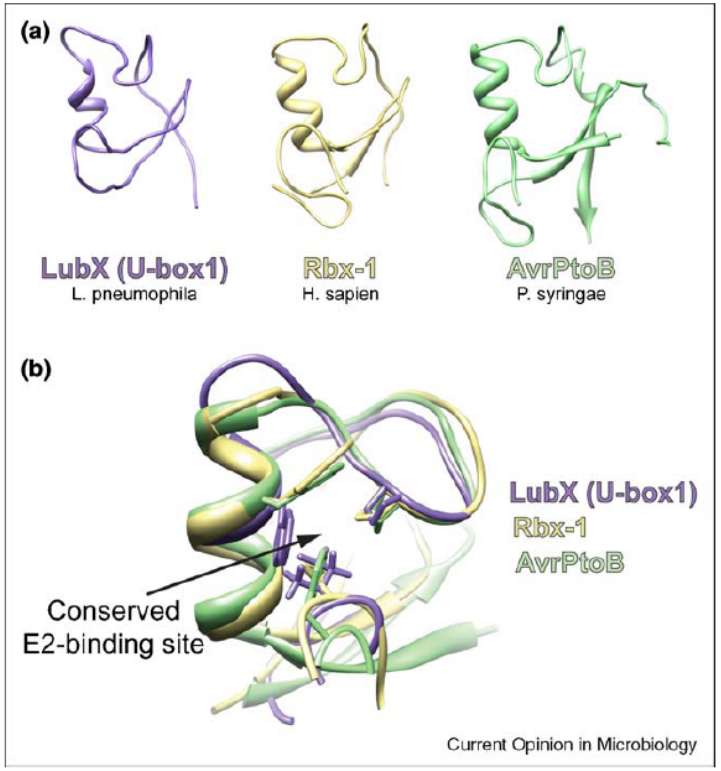
- Do not encode for ubiquitin, yet still have an N-end rule pathway.
- In *E. coli*, Arg and Lys are the secondary destabilizing residues, in contrast to being primary destabilizing in eukaryotes.

  To these, lysine or phenylalanine are conjugated to form a primary signal for degradation by the proteosome-like ClpAP. N-terminal W and Y can also act as destabilizing factors.

- Leucine-conjugating transferases appear to be confined to bacteria; and arginine-conjugating transferases to eukaryotes.



Some pathogenic bacteria have evolved proteins that appear to be molecular mimics of eukaryotic E3 ligases, enabling them to commandeer various aspects of the machinery of host cells (Hicks and Galan, 2010)



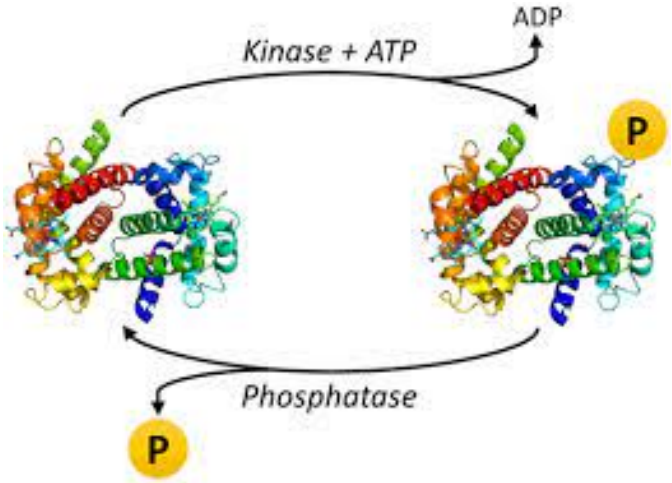
Bacterial mimics of eukaryotic RING/U-box E3 ligases. **(a)** Using the PHYRE threading program, the sequence of U-box1 of *L. pneumophila* LubX was aligned to known structures and the structure was modeled to its best fit, human E3 traf6 (*E*-value of $2.6e^{-11}$; estimated precision of 100%); the RING/U-box structure of *H. sapiens*, Rbx-1 (PDB ID 3DPL); the core fold of *P. syringae*, AvrPtoB (PDB ID 2FD4). **(b)** Visualization of the E2-binding site residues of Rbx-1 with homologous regions in LubX and AvrPtoB. The three putative E2-binding residues are shown.

Post-translational Modification: phosphorylation.

- Kinases add phosphoryl groups to specific AAs, usually serine and threonine residues, but also to tyrosine in animals, and arginine, aspartic acid, cysteine, and histidine in bacteria.

- Recognition sites are simple, generally consisting of the substrate amino acid plus just 2 to 4 flanking AAs.

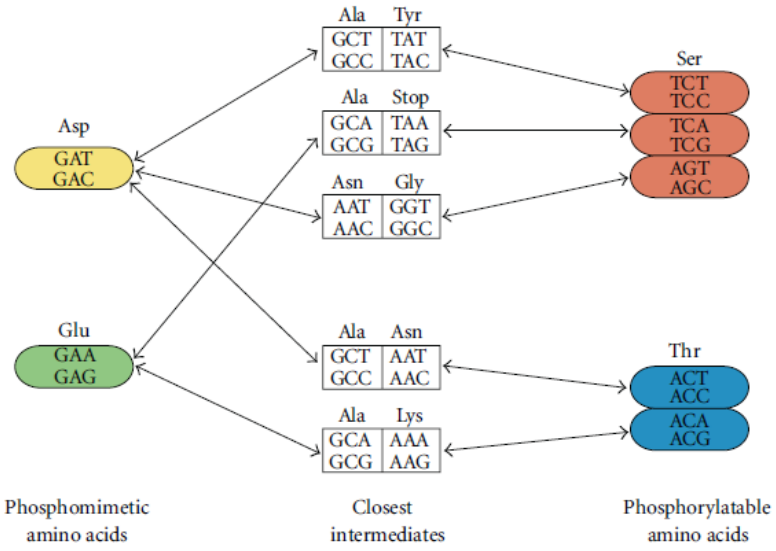
- Raises opportunities for substantial promiscuity.



- Widely deployed:
  - 2% of the yeast genome is devoted to protein kinases, with ~10,000 phosphosites distributed over ~5,000 proteins.
  - Over 500 kinases and 200 phosphatases are encoded in the human genome, and there are >150,000 phosphosites distributed over ~17,000 proteins.

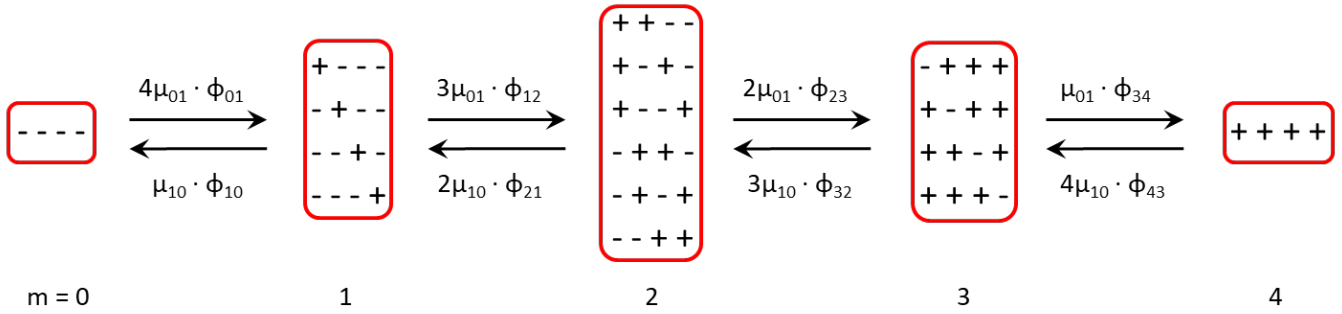
Evolutionary Meandering of Phosphosites

- Phosphosites tend to be clustered on protein surfaces, with selection favoring particular total levels of negative charge.
- The precise locations of individual sites appear free to wander, conditional on the maintenance of total charge.
- Glutamic acid and aspartic acid, both negatively charged, are phosphomimetic, and may transition to conventional phosphosites, although this requires two mutations.
- Emerging picture: the degree of a protein's phosphorylation status is essentially a quantitative trait under stabilizing selection for an appropriate total positive charge, with the specific locations of the affected residues free to wander in an effectively neutral fashion.

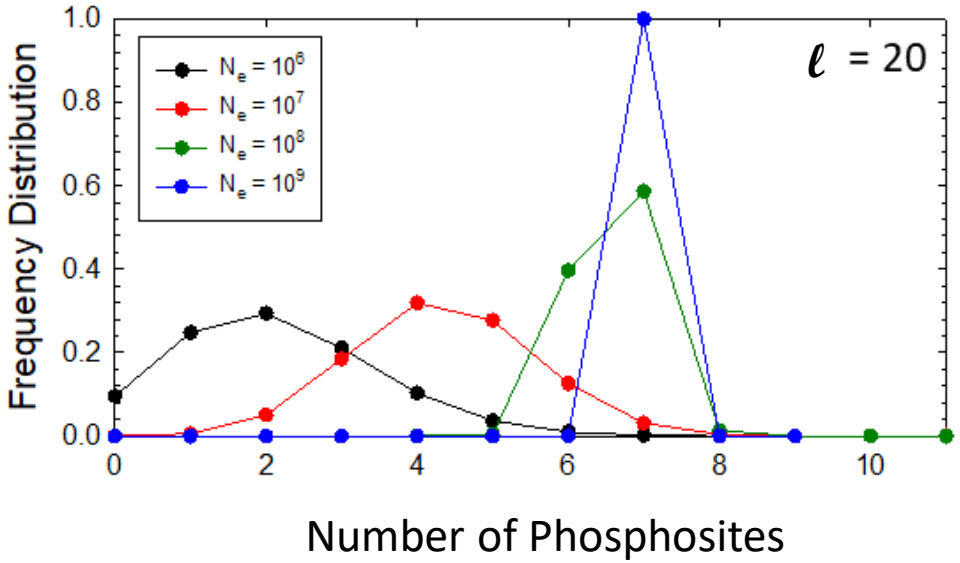


Diss et al. (2012)

Long-term Steady-State Evolutionary Distribution of the Underpinnings of a Quantitative Trait



- Transitions between alternative states driven by mutation, selection, and drift.



- Equilibrium (temporal) distribution.

General Evolutionary Principles / Challenges Illustrated by Protein-Management Systems

- Evolution (and reversals) of multimeric structures and degree of heteromery: chaperones.

- Evolution of molecular languages: chaperone / client-protein recognition; N-end rule in the ubiquitin system; phosphosite recognition by kinases.

- Exploitation by the use of molecular mimics in pathogens: hijacking chaperones and the ubiquitin system.

- Effectively neutral evolution of the underlying components of traits under stabilizing selection: phosphorylation.